SMU Data Science Review

Volume 1 | Number 1 Article 12

2018

Case Study: Using Crime Data and Open Source Data to Design a Police Patrol Area

Brent Allen
Southern Methodist University, brenta@smu.edu

Case Study: Using Crime Data and Open Source Data to Design a Police Patrol Area

Brent Allen*

Southern Methodist University
Masters of Science in Data Science
Dallas, TX 75275

Abstract. This case study examines how to use existing crime data augmented with open source data to design a patrol area. We used the a demand signal of "calls for service" vice reports which summarize calls for service. Additionally, we augmented our existing data with traffic data from Google Maps. Traffic delays did not correspond to traffic incidents reported in the area examined. These data were plotted geographically to aid in the determination of the new patrol area. The new patrol area was created around natural geographic boundaries, the density of calls for service and police operational experience.

Keywords: Police, Patrol Area, Data Visualization, Geographic Plots

1 Introduction

Many police departments are encountering hurdles determining how to use crime data and open source data to prevent crime. In local government decision making is decentralized with respect to utilizing existing open source technology in the government systems. Security concerns and the general lack of familiarity with analytics make decision makers nervous. Additionally, government institutions tend to be late adopters to new techniques. Analyzing data and presenting information is an area where local government can greatly improve operations and decision making. The team used a combination of open source algorithms, crime and traffic data to recommend a new patrol area for the police department. This project allows a unique opportunity to take a data driven approach in creating a new patrol area and by extension keeping our community safe.

A police patrol area is a geographic region of a city in which one or more police units are assigned. The patrol units are not confined to this area. Typically, these geographic regions are created based on the needs of the community and operational capabilities of the police department.

This case study uses crime data from the city of Irving, Texas. The Irving Police Department segments the city into 23 patrol areas knowns as beats (Figure

* Brent Allen is a Data Scientist with Flowserve Commercial Operations. This work was performed as part of his Master of Science in Data Science *Above and Beyond* coursework.

1). These beats are used to organize the police force, efficiently operationalize the force and collect data for analysis. The existing beats were created 12 years ago and the city has changed significantly since then. The police department used the expertise of the command staff, patrol sergeants and historical examples to design the current patrol areas.

The team wanted to take a data-driven approached utilizing the crime data available and open source data. The team also wanted to let the data tell the story and minimize any pre-conceived notions that may skew the shape of the new beat. We were given a search area (Figure 2) where the new beat will be created. This area is a combination of beats 53 and 51 which encompasses parts of Las Colinas, southern Valley Ranch and Irving. This search area is between latitude / longitude 32.895787, -96.950260 and 32.845798, -96.911913.

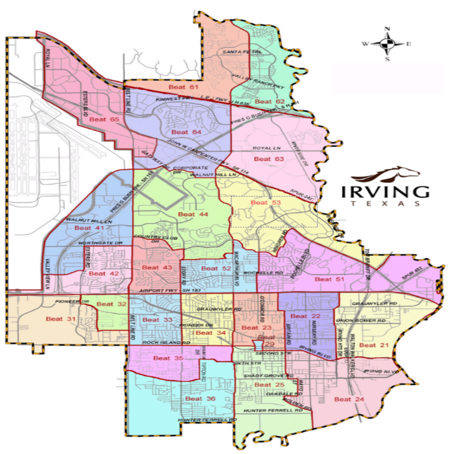


Fig. 1. current patrol areas

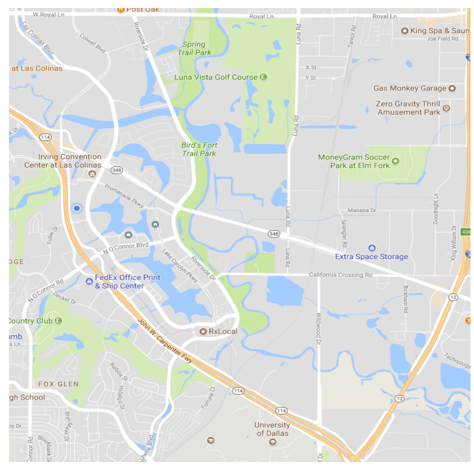


Fig. 2. search area for the new patrol area

In order to determine how the new patrol area will be developed some decision criteria and metrics are evaluated. Several people from different parts of the police department were invited to determine the metrics that will be used to evaluate the new patrol area. A Six Sigma / LEAN decision making technique, dot voting, was used to determine the metrics for this case study. The members included a police Captain which brought a "boots on the ground" and operational expertise point of view, an geographic analyst whom has access to the current and future construction projects planned within the city limits, a dispatch operator which understands how the communication systems operate and a database administrator that understands how the systems transfer data.

The metrics we will use are the type of criminal offense, the number of offenses, the number of calls for service, the priority of the call and traffic delays. The type of offense is a categorical variable which describes the incident. When a call is reported to the dispatcher through the 911 system this category is selected from a drop-down menu.

The crime density is number of offenses within a geographic region. Crime density changes with respect to time and the type of offense. This metric will not be the number of incidents per square mile. The crime density is a visual metric. These data will be plotted on a map and displayed based on the type of crime and the priority of the call. The latitude and longitude for each criminal incident are captured in the database. The location is updated as the offense changes location.

The number of calls for service are the calls that are received by the 911 system and sent to a Patrol Officer to respond. The call is recorded in the database with a key value associated with the call. The number of calls will vary with respect the type of offense, priority, date-time and geographic location.

The call priority is an ordinal variable that represents the severity of the incident. This metric ranges from 1 to 9 where 1 is the most severe.

The time of the traffic delay is the difference of the travel time with no traffic compared to current traffic conditions. This data is collected through the Google Maps Application Programing Interface (API). The address, travel times and delay times are available through the API. The team defined fourteen routes through the search area to capture the traffic patterns. These traffic patterns cover the major streets through the search area. The application collects the data every 7 minutes and stores that information in a local Microsoft SQL database. The date and time of the traffic query is collected to allow date-time analysis of the traffic delays.

The police officer time breakdown was originally included in the metrics considered to evaluate a new beat. The metric consisted of the ratios of officers time allocation while on patrol. An officer is available to respond to an incident from reporting on duty to when they leave their shift. The time is broken into several categories; time available to respond to an incident, the response time of an officer to an incident and the time to conduct the call for service.

The team wanted to include weather conditions in the analysis. The daily temperature, precipitation, wind speed and direction and any storm or catastrophic event.

The remainder of this paper is organized as follows. In Section 2 we review the basic design and purpose of patrol areas. In Section 3 the criteria for optimization is discussed. In Section 4 we examine the data sources used in our analysis. In Section 5 the pre-processing steps and issues are discussed. In Section 6 the analysis of the data is presented. In Section 7 the new patrol area recommendation is presented. In Section 8 we conclude with a recommendation of how to reproduce this case study.

2 Patrol Area Operations

While on patrol a police officer responds to calls for service, observes activity in and round their current position and available if a citizen needs assistance. The

Patrol Officers are dispatched to a patrol area, "beat", during their shift. In the briefing before each shift different points of reference and activities to be aware of are pointed out by the Patrol Sergeant. Each Patrol Officer is functionally responsible to the Sector Sergeant during their time on patrol. During a shift, a patrol officer travels the area assigned vigilant to areas of known activity and identifying suspicious persons or incidents. A Patrol Officer is responsible for protecting the lives and property of the residents and visitors of the City of Irving, knowledge of potential problem areas and criminal elements in the assigned patrol area, recognizing offense patterns and probabilities, understand the geographic characteristics of the assigned patrol area, interacting with the citizenry in a positive manner, communicating with the Sector Sergeant if unusual activity is observed, answering calls for service in a prompt and professional manner, conducting investigations in accordance with the Standard Operating Procedure (SOP), promoting the safe and orderly flow of traffic in the assigned Patrol Area and reporting all engineering defects (streets, traffic flow, etc...). A Patrol Officer may be required to respond to a "back-up" request from another officer. In this case the Patrol Officer will may leave his/her patrol area to respond to the backup request. If an officer observes an incident they will respond accordingly to their training and experience to handle the situation. Each incident is reported through the 911 dispatch system and is recorded in the Computer-Aided Dispatch (CAD) system.

3 Patrol Area Optimization

For the purposes of this case study optimization is achieved by looking at the incident density by type and priority over the past year. We want a patrol area to be able to be serviced by a patrol officer(s) and respond to 95% of the reported incidents on the shift within a reasonable time. In this case a reasonable time depends on the type and priority of the incident. Calls for service with a higher priority (lower number as defined in Section 4) are required to respond quicker then calls with a lower priority (higher number as defined in Section 4). The response time for each call was not calculated due to the complexity of the source system. As a result the "reasonableness" of a response time will be measured by the size and traffic pattern of the patrol area. The smaller the patrol area the faster an officer will be able to respond. The lack of traffic delays in a patrol area the faster an officer will be able to respond.

4 Data Sources

The crime data is contained in two relational databases; Records Management System (RMS) and Computer-Aided Dispatch (CAD). RMS holds all the records related to police and crime activity collected by the police department. The patrol areas in Figure 1 were designed by from reports held in RMS. The data in RMS includes people, locations, incidents and narratives in a highly relational database. RMS contains only reports. Not every call for service will result in a report. The CAD system holds the records that the police respond to and a status of the response. In using the CAD database we are able to take advantage of the true demand signal of the community. We made the decision to assume all the data in the criminal record database is correct. There is no indication that the data in systems is false or manipulated.

The data in RMS is organized into several tables. The "incident table" holds the records of the incidents that officers report. These records are summarized and do not show any changes in the dimensions examined. We decided to focus on the CAD tables which show all the calls for service.

The data in CAD is separated into three tables that are constantly updated. The three tables are identified as the "segment" tables SEGMENT_01, SEGMENT_02, SEGMENT_03. Table SEGMENT_01 contains the time information, properties of the call such as the call type while Table SEGMENT_02 holds miscellaneous information and Table SEGMENT_03 holds geographic information. Each segment table has a dimension which is an index for each record key.

Table 1. Segment Table 01

column name	column type	column description
CALLKEY	alpha-numeric string	the call identifier, unique to only one call through the 911 dispatch system
SEGMENT_SEQ_CTRL	integer	the sequence of the call identifier, this number identifies the order of the records associated with the call, ranges from 1 to max(n)
SEGMENT_DATE	date-time	the date of the record to the day, this column doesn't include the time values; e.g. 2017-09-01 00:00:00
SEGMENT_CALL_TYPE	alpha-numeric string	the category or type of incident reported through the 911 dispatch system, some examples are '911', 'THAZ', 'NOISE',...
SEGMENT_TIME	numeric string	the time in 24 hour format but is in text and not formated as a time variable in the database; e.g. '094941' is 09:49:01 on the day of the reocrd
SEGMENT_INT	integer	time in milliseconds since the origin time, January 01, 1970
SEGMENT_PRIORITY	numeric string	the priority of the call through the 911 dispatch system, ranges from 1 - 9 where 1 is the most important and 9 is the least important

The data descriptions in Table SEGMENT_01 are in Table 1. The CALLKEY field is the primary key for all the segment tables. This key is not the used like a SQL primary key. This is a reference repeated multiple times and is unique to only the incident referenced. The SEGMENT_SEQ_CTRL is the number that ensures the records in each table are synchronized. For each CALLKEY the three segment tables have the same number of rows. A significant amount of the fields are NULL since only a few fields are updated for each entry. The SEGMENT_SEQ_CTRL number keeps all the entries in the order entered into the system. There are three fields that capture a time or date. The SEGMENT_DATE field captures the date as a date-time format. The field SEGMENT_TIME is a string that records the time of the entry in 24-hour format. Since this is a string a date-time conversions needs to be conducted to allow this field to be computationally useful. The field SEGMENT_INT is a number which captures the time in milliseconds for the record entry. This time is the number of milliseconds from the origin date. This field can be used to calculate differences in time of records for further study. The field SEGMENT_CALL_TYPE is a categorical field which identifies the type of criminal activity or police activity conducted.

An example of a record in the incident table is Table 2.

Table 2. Example for first segment table

column name	example
CALLKEY	CK0001
SEGMENT_SEQ_CTRL	1
SEGMENT_DATE	2017-09-01 00:00:00
SEGMENT_CALL_TYPE	911
SEGMENT_TIME	094941
SEGMENT_INT	1534786598
SEGMENT_PRIORITY	2

Table 3. Third segment table in CAD

column name	column type	column description
CALLKEY	alpha-numeric string	the call identifier, unique to only one call through the 911 dispatch system
SEGMENT_SEQ_CTRL	integer	the sequence of the call identifier, this number identifies the order of the records associated with the call, ranges from 1 to max(n)
SEGMENT_GEOXCOORD	numeric string	the longitude of the report as reported through 911 dispatch system
SEGMENT_GEOYCOORD	numeric string	the latitude of the report as reported through the 911 dispatch system

The data in Table SEGMENT_03 relates to the responding officer and latitude / longitude as well as any updates in location of the reported incident. The primary fields for this table are in Table 3. The CALLKEY is the same reference as Table 1, the primary reference for the table and the incident reported through the dispatch system. The SEGMENT_SEQ_CTRL number is the same as Table 1 which keeps the order of the entries consistent in each of the three tables. The geographic coordinates (geo-codes) are in latitude and longitude format as floats. The latitude is the Y-Coordinate but is reported first. North of the equator are positive numbers and south of the equator are negative numbers. This numbers range from 0.0 to 90.0 or -90.0. Longitude is the X-Coordinate and is reported second. These numbers range from 0.0 to 180.0 and -180.0. West of the Prime Meridian (Greenwich, England) are negative numbers and positive numbers are East of the Prime Meridian. Geo-coords 32.86, -96.93 is 32.86 degrees north of the equator and 96.93 degrees west of the Prime Meridian.

The example of geo-codes is in Table 4.

Table 4. Example for third segment table

column name	example
CALLKEY	CK0001
SEGMENT_SEQ_CTRL	2
SEGMENT_GEOXCOORD	-96.931234
SEGMENT_GEOYCOORD	32.861234

The traffic data is collected from the Google Maps API and stored in a local SQL database. Each route covers the major surface streets where new patrol area will be created. The routes do not cover highways and interstates.

The Table 5 contains the information to query the Google Maps API and obtain the information about the surface street traffic. The field string_route_id and the string_route_desc identify and describe the route that will be measured. Each route can contain multiple legs where the traffic delay is measured. The fields float_route_lat_start, float_route_lng_start, float_route_lat_stop and float_route_lng_stop identify the start and stop points of the route. The field int_leg_num identifies the sequence of the legs of the route. The fields string_leg_addr_start and string_leg_addr_stop are the beginning and the end addresses of the route leg. The fields float_leg_lat_start, float_leg_lng_start, float_leg_lat_stop and float_leg_lng_stop are the latitude and longitude coordinates of the beginning and end of the route leg.

Table 6 is an example of the route data.

Based on the geography and the street design of the area we broke the area into 14 routes with multiple segments. Initially we tried to only define routes with multiple way-points using the Polyline Algorithm[1]. The problem is the traffic times were not returned. That not only impacted the implementation of obtaining the traffic delays but also the number of queries that could be conducted.

[1] url: https://developers.google.com/maps/documentation/utilities/polylinealgorithm

Table 5. Routes Table

column name	column type	column description
string_route_id	alpha string	the route identifier
string_route_desc	alpha string	the description of the route
float_route_lat_start	float	the latitude for the start of the route
float_route_lng_start	float	the longitude for the start of the route
float_route_lat_stop	float	the latitude for the stop of the route
float_route_lng_stop	float	the longitude for the stop of the route
int_leg_num	integer	the leg number for the route; each route has multiple legs
string_leg_addr_start	alpha-numeric string	the address for the start of the leg of the route
string_leg_addr_stop	alpha-numeric string	the address for the end of the leg of the route
float_leg_lat_start	float	the latitude for the start of the leg
float_leg_lng_start	float	the longitude for the start of the leg
float_leg_lat_stop	float	the latitude for the stop of the leg
float_leg_lng_stop	float	the longitude for the stop of the leg

Based on Google Maps rates[2] the first 1,000 are free with only an API key. A credit card is needed to verify the identity of the person using the API. Then the limit is increased to 150,000 requests per day. Using the polyline implementation of a route in Google Maps we could cover the 14 routes with the 1,000 requests per day which would limit frequency to about 20 minutes. The number of queries increased to 66 vice 14. We had to break up the routes into smaller segments. As a result we needed to increase the number of queries to 150,000. This allowed the team to query Google Maps every 7 minutes. The python script is running continuously for about a 2 months at the time of publishing. There is approximately 750k records in the traffic database which cover every major surface street for the new patrol area.

The results of the Google Maps API query are in Table 7. Most of the fields are the same as in Table 5 except for the following fields. The field date_calculation is the date-time of the query to the Google Maps API. The field string_day is the day of the week based on the field date_calculation. The field string_leg_time_text is the text description of the length of the route leg

[2] url: https://developers.google.com/maps/pricing-and-plans/

Table 6. Example of one record in the route table

column name	example
string_route_id	route_01
string_route_desc	frontage road 114 northwest
float_route_lat_start	32.
float_route_lng_start	-96.
float_route_lat_stop	32.
float_route_lng_stop	-96.
int_leg_num	0
string_leg_addr_start	1111 Northwest highway, Irving Texas, 75039
string_leg_addr_stop	1112 Northwest highway, Irving Texas, 75039
float_leg_lat_start	32..
float_leg_lng_start	-96.
float_leg_lat_stop	32.
float_leg_lng_stop	-96.

(e.g. 0.5 miles). The field int_leg_dist_yds is the distance of the route leg in yards as an integer. The field string_leg_time_text is the time as a text field returned from the Google Maps API. The field int_leg_time_seconds is the time on average in which a car will travel the route leg under normal conditions. The field string_leg_traffic_text is the text string of how long the traffic delay is from the Google Maps API. The field int_leg_traffic_seconds is the time in seconds for the traffic delay.

An example of the traffic results is in Table 8.

5 Pre-Processing

The two primary tables used to gather data for new patrol area are the INCIDENT and CALLS FOR SERVICE tables in the CAD database. The INCIDENT tables catalog any incident that a policeman responds to. The CALLS FOR SERVICE tables record any phone call to report an observation or crime through the 911 system.

Figure 2 shows the area where the data from the CAD system will be focused and gathered. The area is bound by the natural barriers of Lake Carolyn, Elm Fork Trinity River and basins, Texas Highways 114 and 12, The Las Colinas Country Club and TPC Four Seasons Las Colinas. The area in Figure 2 is between latitude 32.895787 and 32.845798 , longitude -96.950260 and -96.911913.

From these tables we collected the necessary fields to build a picture of the activity in this area. We segmented the data into three main groups; geographic, call type and call priority. All the records are from 06 August 2006 to 12 Sept 2017. September 12th is an arbitrary date that was chooses to allow the completion of the project based on the volume of data.

Table 7. Traffic information table

column name	column type	column description
date_calculation	date-time	the date-time of the query to Google Maps
string_day	alpha string	the day of the week of the query to Google Maps
string_route_id	alpha string	the route identifier
float_route_lat_start	float	the latitude for the start of the route
float_route_lng_start	float	the longitude for the start of the route
float_route_lat_stop	float	the latitude for the stop of the route
float_route_lng_stop	float	the longitude for the stop of the route
int_leg_num	integer	the leg number for the route; each route has multiple legs
string_leg_dist_text	alpha-numeric string	the distance of the leg from Google Maps in text
int_leg_dist_yds	integer	the distance of the leg from Google Maps in yards
string_leg_time_text	alpha-numeric string	the time to travel the leg from Google Maps under normal conditions with no traffic
int_leg_time_seconds	integer	the time to travel the leg from Google Maps in seconds under normal conditions with no traffic
string_leg_traffic_text	alpha-numeric string	the time to travel the leg under current traffic conditions from Google Maps
int_leg_traffic_seconds	integer	the time to travel the leg under current traffic conditions from Google Maps
string_leg_addr_start	alpha-numeric string	the address for the start of the leg of the route
string_leg_addr_stop	alpha-numeric string	the address for the end of the leg of the route
float_leg_lat_start	float	the latitude for the start of the leg
float_leg_lng_start	float	the longitude for the start of the leg
float_leg_lat_stop	float	the latitude for the stop of the leg
float_leg_lng_stop	float	the longitude for the stop of the leg

The data collected in the two segment tables was combined on the key in each table to produce 469,285 total records. The number of distinct cases, as identified by unique key values, is 66,556. There is roughly 6 or 7 records for each criminal incident in the database.

As in most data-munging / pre-processing techniques there are decisions that need to be made in order to bring clarity of the data. In this case free text fields produce an issue with spelling. The call type description seems to be a free text field based on the number of phrases spelled differently. The phrase 2x is also spelled 2 X. The phrase Hang Up is spelled Hng Up. The phrase Aud/intrusion is spelled Audintrusion and Audintrusion. There are several ways to attack this problem. The algorithmic method is to break each phrase into a count vector and determine the cosine distance between each phrase. The duplicates are determined by a clustering technique (Density Clustering may be an option) or utilize the expertise of the team from the Irving Police Department. The advisory team provided additional insight into the meaning of not only the call type but also the description of the call type. Several call type categories were able to be combined due to similarity of the offense and changes in recording procedure and call tracking since 2006. The team examined the data and this

Table 8. Example of one record in the traffic information table

column name	example
date_calculation	2017-09-01 05:42:32
string_day	Friday
string_route_id	route_01
string_route_desc	frontage road 114 northwest
float_route_lat_start	32.
float_route_lng_start	-96.
float_route_lat_stop	32.
float_route_lng_stop	-96.
int_leg_num	0
string_leg_dist_text	0.3 mi
int_leg_dist_yds	528
string_leg_time_text	.7 min
int_leg_time_seconds	42
string_leg_traffic_text	1 min
int_leg_traffic_seconds	65
string_leg_addr_start	1111 Northwest highway, Irving Texas, 75039
string_leg_addr_stop	1112 Northwest highway, Irving Texas, 75039
float_leg_lat_start	32.
float_leg_lng_start	-96.
float_leg_lat_stop	32.
float_leg_lng_stop	-96.

data is a free text field and decided not to include this field in the analysis. In the future natural language processing could be used to glean some additional information from this field.

The data was combined in a summary table from all the raw data tables except for the traffic data. Each table is combined on the record key and the record index. Table 9 is the analysis table with the combined data.

During this phase of the project two metrics were taken out of the potential evaluation list. The Officer time breakdown metric was not able to be determined effectively due to the lack of specificity of the data. The goal of this metric is to look at the amount of time an officer spends on a call by call type. Due to the lack of a dimension which identifies an officer on each call we were not able to define this metric within the context of the data. The data can relate a unit (car, motorcycle, etc.) to a record but not an officer.

The weather metric was also dropped from the decision criteria. The team determine from experience and intuition that temperature, frequency of a rain and the number of weather changes effect criminal activity. In the timespan for this project we were not able to obtain weather data in a form that is usable.

6 Data Analysis

The area being analyzed is changing rapidly. New work/live/play areas are being built, construction is occurring on many surface roads and highways and business are opening or expanding. Calls for service are affected by these changes. The traditional metric for calculating the number of patrol officers for a community

Table 9. Analysis Table

column name	column type	column description
string_callkey	alpha-numeric string	the call identifier, unique to only one call through the 911 dispatch system
int_segment_num	integer	the sequence of the call identifier, this number identifies the order of the records associated with the call, ranges from 1 to max(n)
date_record	date-time	the date of the record to the day, this column doesn't include the time values; e.g. 2017-09-01 00:00:00
string_time	numeric string	the time in 24 hour format but is in text and not formated as a time variable in the database; e.g. '094941' is 09:49:01 on the day of the reocrd
int_time	integer	time in milliseconds since the origin time, January 01, 1970
float_geoxcoord	float	the longitude of the report as reported through 911 dispatch system
float_geoycoord	float	the latitude of the report as reported through the 911 dispatch system
string_callpriority	numeric string	the priority of the call through the 911 dispatch system, ranges from 1 - 9 where 1 is the most important and 9 is the least important
string_calltype	alpha-numeric string	the category or type of incident reported through the 911 dispatch system, some examples are '911', 'THAZ', 'NOISE',...

is a certain number per thousand of the population. That metric has come into question over the recent years.

We initially looked at the data for the past 12 months to understand how the calls for service categorize from the call priority and call type. Figures 3 and 4 show the basic breakdown of the call priorities and call types for the search area. Call priority 2 was significantly higher then next call priority which was 9. In speaking with our team most of the calls that are not classified with a call type are prioritized at a 9 until further clarification is entered. Call priority 1, the most emergent priority, was the lowest percentage of the calls for service. The call type showed that vehicle tows (TOW) and 911 calls (911) where very close in percentage of calls for service. Traffic related calls, traffic hazard (THAZ) and accidents (ACC), were low.

We wanted to see where these incidents were geographically. Figures 5 and 6 show the calls geographically by call priority and type. The plots show there is significant activity on Texas Highway 114. Some of the priority 2 and 6 calls are along the highway. Although Towing and 911 do not seem to show up as much as Alarms and Disturbances. Alarms and Disturbances occur in the same places as the towing incidents and the Alarm and Disturbance calls are plotted last in the plot sequence. The density of the incidents is telling. There seems to be three distinct areas, excluding TX Highway 114, the bottom left corner (area BL), the middle third (area m3) and the top of the search area (area top). Area BL is

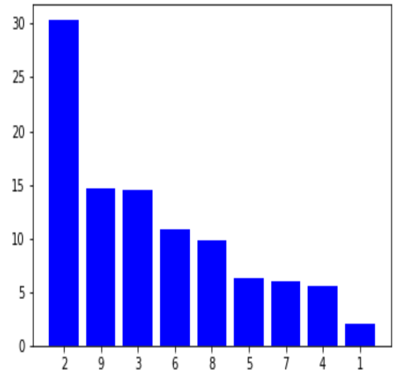


Fig. 3. call priority over the past year by percentage

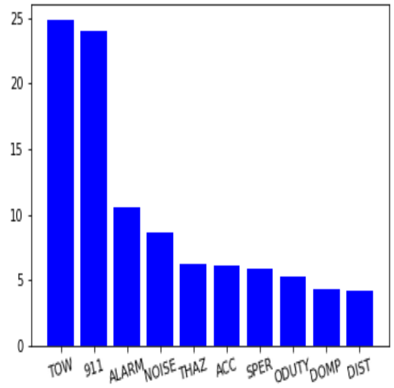


Fig. 4. call type over the past year by percentage

separated from area m3 by the Las Colinas Country Club and area BL has a high proportion of single family homes and some commercial real estate. Area m3 contains a mix of commercial real estate which contains international businesses, condos and apartment complexes, and some restaurants. A new entertainment complex is recently completed, the Toyota Music Factory. This facility hosts concerts, comedy shows and other celebrity appearances. Additionally, the Irving Convention Center is in this area and two Dallas Area Rail Transit (DART) stations are in this area. The top area is a mix of apartment complexes and some commercial real estate.

We needed to understand how the call priorities and call types differentiated geographically. In Figures 6, 7, 11 and 12 we can see several patterns emerging. TX Highway 114 produces a lot of priority 2 calls for service which seem to be disproportionately 911. 911 category calls do not seem to be in residential areas compared to other category 2 calls. Towing calls seem to be mostly around apartments, condos and commercial businesses. The geographic pattern of the calls remain the same with the three distinct areas excluding TX Highway 114. The combination of Figures 7, 8, 9 and 10 show that 911 calls are the most significant by count and percentage for priority 2 calls in the search area. These calls plot mostly along TX Highway 114 and major surface streets and not as significant in residential areas and apartments. In contrast calls for tows (TOW) are the highest call type as shown in Figures 8 and 4 which are in the same vicinity as apartment complexes, condos and businesses.

In Figures 11, 13 and 14 show the calls for service for call priority 3. Alarm (ALARM) and suspicious person (SPER) calls make up the majority of the calls for service for priority 3 calls. Alarm calls are mostly in residential areas while noise calls are in business areas as well as apartment complexes as indicated in

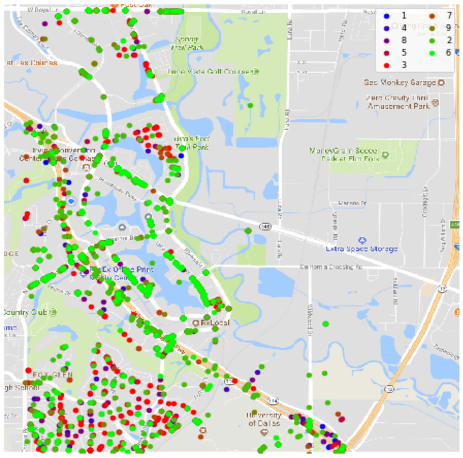


Fig. 5. call priority over the last year

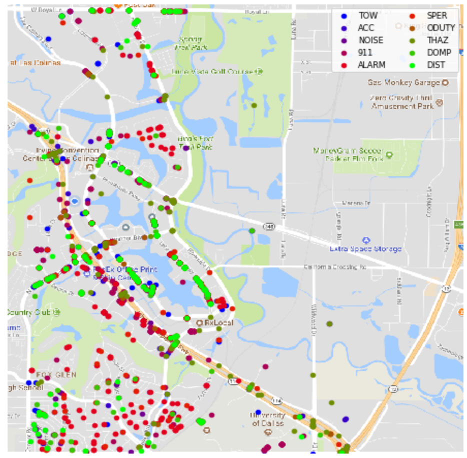


Fig. 6. top 10 call types over the last year

Figure 12. The density pattern is similar to the three main areas as indicated in the previous Figures 7 and 8.

How does this relate to the traffic delays? Traditionally, traffic patterns are heaviest in areas going to areas of business and education in the morning and traveling away from those areas in the evening. Does that pattern exists in this area and if so how does it affect calls for service?

Figures 17 and 18 show the top 20 median and mean traffic delays by day of the week, hour and route. Figures 15 and 16 show the location of the traffic delays and the relative location to the top 10 call types in the area. The size of the semi-transparent circle relates to the number of delays. The largest and most frequent delays is the purple circle where two major Texas highways meet and encompasses highway entrances. All the top 20 time delays are in the afternoon, 4pm through 7pm. Most people are traveling home from work at this time. All these traffic delay locations are entrances to Texas highways. This traffic delay information does not include the major highways only surface streets.

Examining the mean and median by hour and day of the week we can glean one interesting piece of information. The average traffic delay is lower the median traffic delay in most cases. The longest median delay is 150 seconds, about two and a half minutes while the highest mean delay is 130 seconds, slightly over two minutes. The last quarter of the top 20 delays is where the mean is higher than the median. Since the mean and median traffic delays by hour and day of the week are low and the median is often higher than the mean a lot of the traffic delays are not that significant in this area.

Traffic delays will affect patrol officers responding to calls for service when they have to transit over TX Highway 114. There are only two major roads that cross TX Highway 114 in this area and two of the four top 20 traffic delay positions is on one of these crossings. Between the hours of 4pm and 7pm patrol

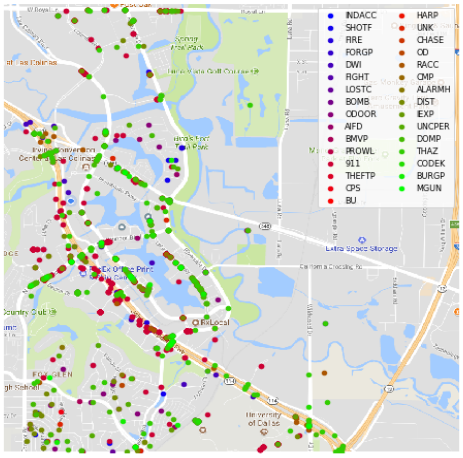


Fig. 7. all call types for priority 2

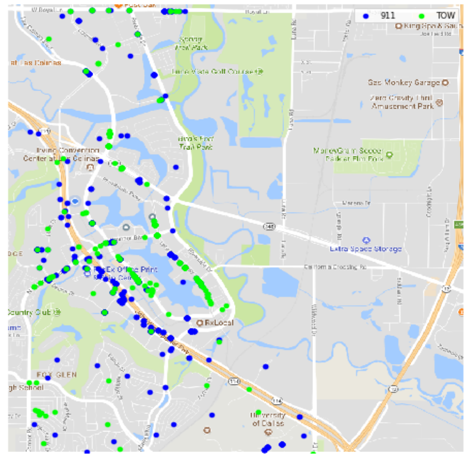


Fig. 8. top 2 call types for priority 2

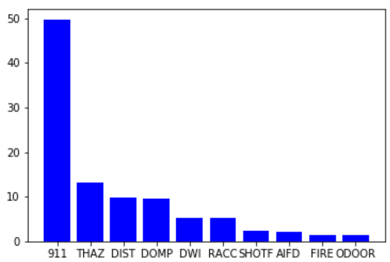


Fig. 9. top 10 call types for priority 2 by percentage

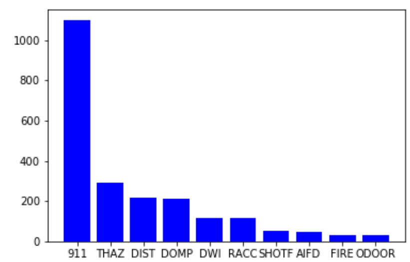


Fig. 10. top 10 call types for priority 2 by count

officers would be able to respond to calls for service quicker if they were pre-position on either side of highway 114 during these hours.

Figures 19 and 20 shows the top 20 traffic delays in relation to traffic hazards and accidents. From a visual inspection there doesn't seem to be a lot of over lap with traffic delays and these calls for service. The orange and the plum colored traffic delays do seem to have some call types in the same vicinity but the green and the purple colored traffic delays do not seem to cover the same amount of traffic incidents. From these visualization traffic delays may not be related to traffic calls for service.

Based on the analysis and Figure 3 through 20 a small area was selected for further analysis. This section is an the area around Lake Carolyn and indicated in Figure 21. We wanted to examine the trends of the data over the past several years. Tables 23 and 24 show the history of a few call priorities and call types in the new search area. There is an upward trend from 2006 to present in most

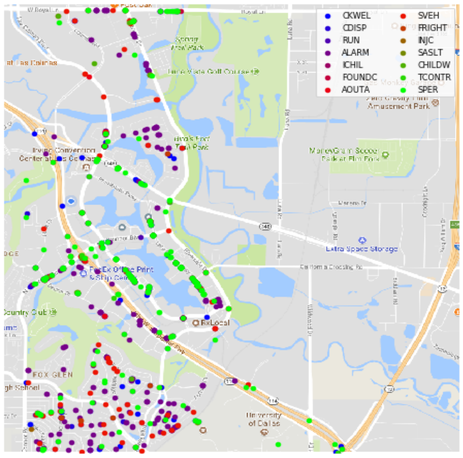


Fig. 11. all call types for priority 3

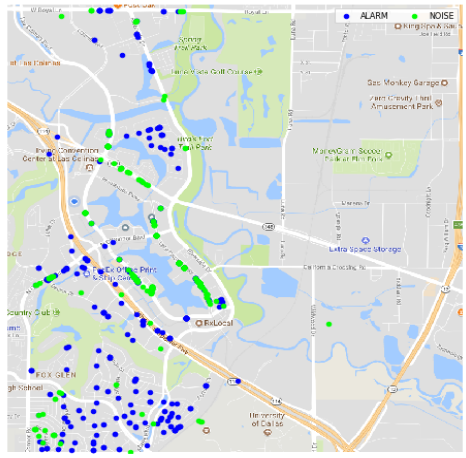


Fig. 12. 3^{rd} and 4^{th} most call types

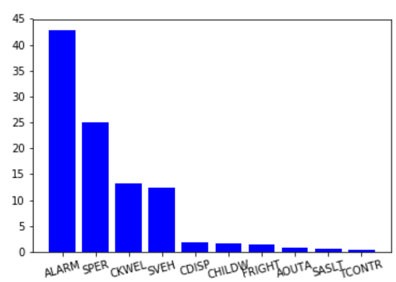


Fig. 13. call priority 3 top 10 call types by percentage

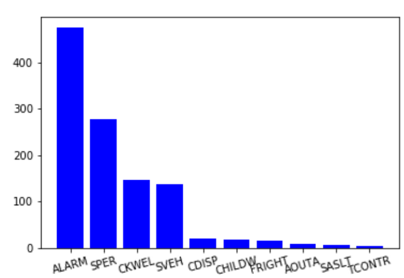


Fig. 14. call priority 3 top 10 call types by count

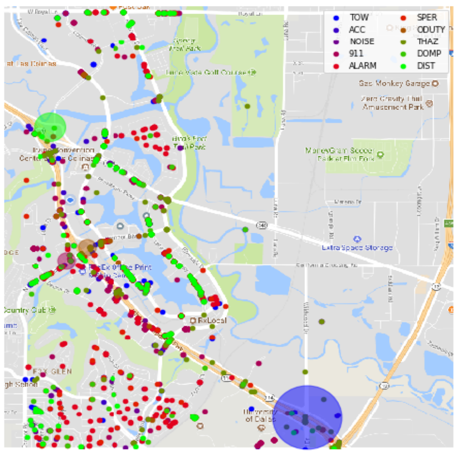


Fig. 15. call types with traffic

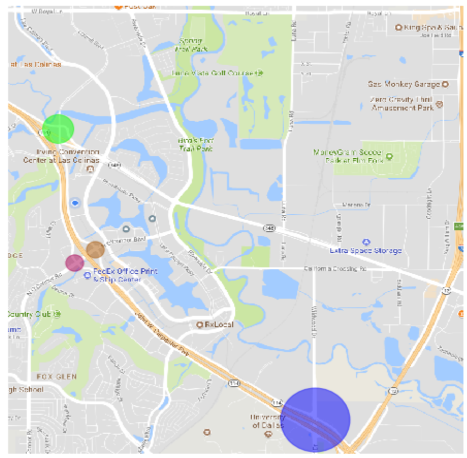


Fig. 16. top 20 traffic delays by day of the week and hour

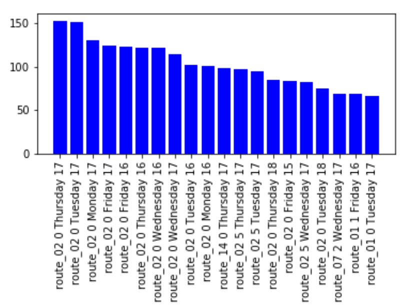


Fig. 17. top 20 traffic delays by day of the week and hour, median

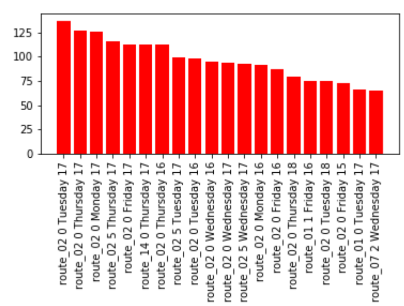


Fig. 18. top 20 traffic delays by day of the week and hour, mean

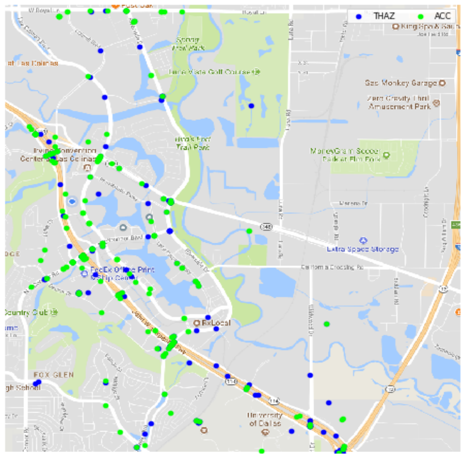


Fig. 19. accidents and traffic hazards

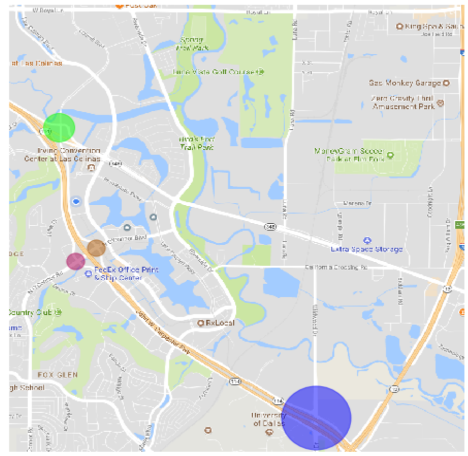


Fig. 20. top 20 traffic delays by day of the week and hour

of the call priorities and call types. 911 calls seem to be oscillate on a 6 year cycle but the data has not been collected long enough to determine that type of trend. Accidents (ACC), disturbances (DIST) and traffic hazards (THAZ) are on an upward trend. Off duty (ODUTY) calls and suspicious persons (SPER) are on a significant and upward trend. This lead the team to determine the calls for this area are changing in proportionality and volume. In general the trend is upward. The top 4 call priorities from the original search area are all on an upward trend.

7 Recommendation

Based on all the analysis presented the recommended patrol area is indicated in Figure 25. This area is bound east of Texas highway 114 and west of Elm Fort Trinity River. The northern boundary is the dike crossed by Las Colinas Blvd and Riverside Dr. The southern boundary is Riverside Drive. The role of a Patrol Officer patrolling TX Highway 114 has not been determined. Based on current practice the officer assigned to this new patrol area would be responsible for covering the western direction of the highway. Further analysis need to be conducted as to the impact of this coverage and the officers time responding to calls for service in the new patrol area.

8 Conclusions

Determining a new patrol area is a complex task. The features considered are often multi-dimensional and not always clear. When starting from scratch the most appropriate option is to bound the problem to a small area and use a

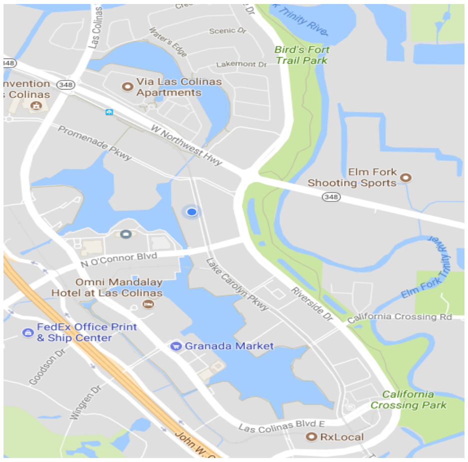


Fig. 21. New search area.

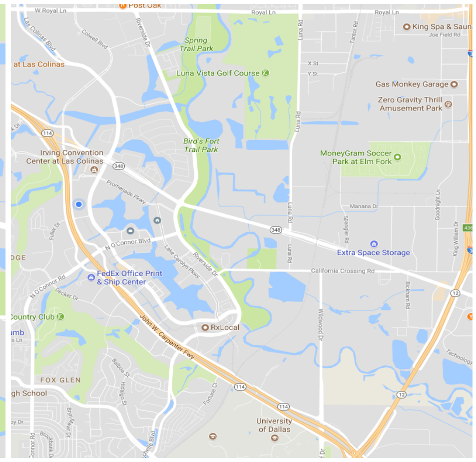


Fig. 22. Original search area.

small subset of features to narrow the focus of the problem. We finalized our metrics to be call priority, call type, density of the calls for service, the traffic delays and geographic boundaries. We did not use the reports filed since not all calls for service are in the reports database. Visualizing the data is key for us. We needed to understand how the calls for service and traffic delays related geographically to call priorities and call types. Allowing the data to lead the team to a conclusion and focusing on the demand allowed us to use a scientific approach and confidently recommend a new patrol area.

Acknowledgments The authors thank the Southern Methodist University Master of Science in Data Science program for the opportunity to perform this work as partial fulfillment of the degree requirements.

The authors would like to thank the Irving Police Department team CAPT John Cullison, SGT Kim Vanderveen, Systems Manager Chris Bryan and Geospacial Manager Quinton Henderson. A special thank you go Chief of Police Jeff Spivey allowing the authors assistance on this project.

Fig. 23. new search area call priorities

call priority call type

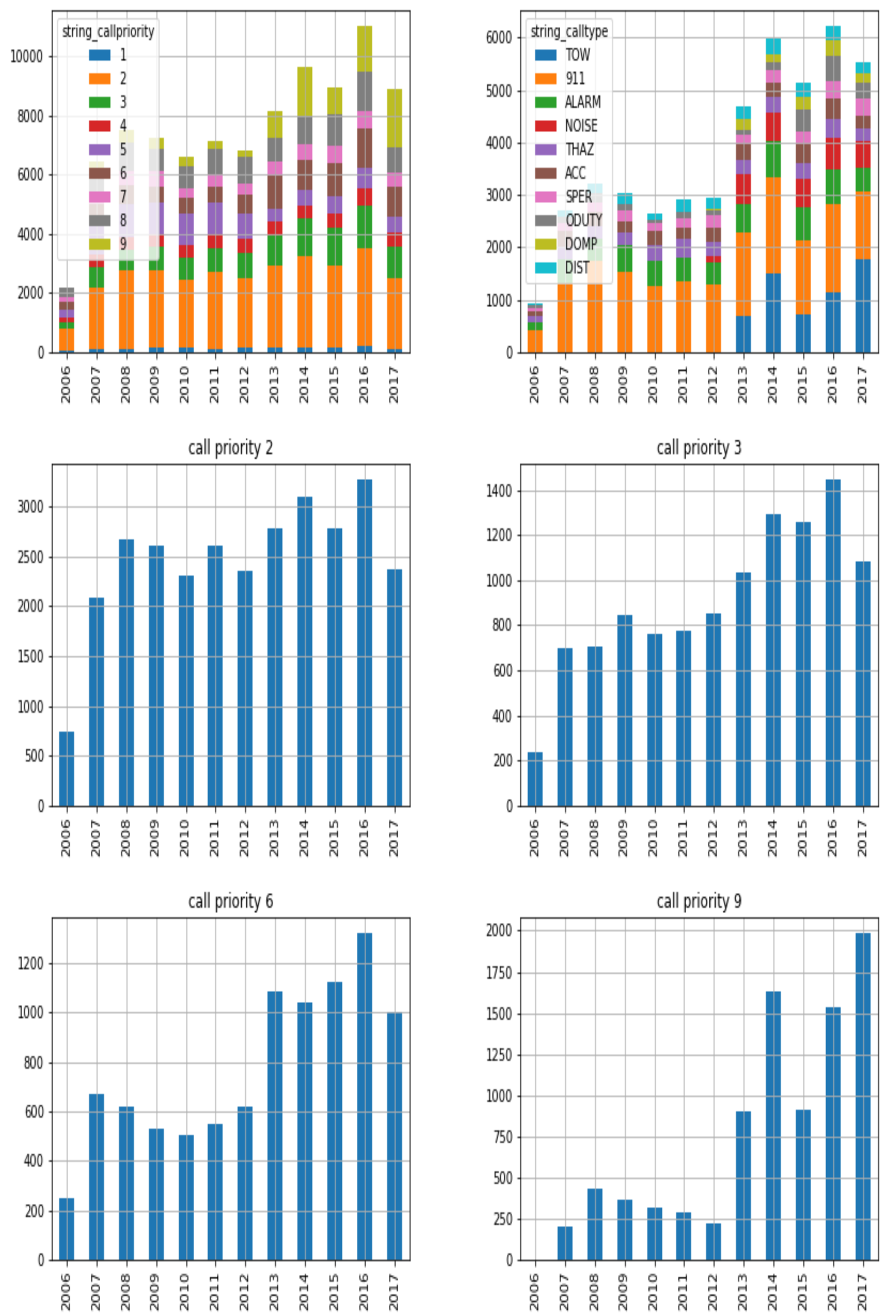
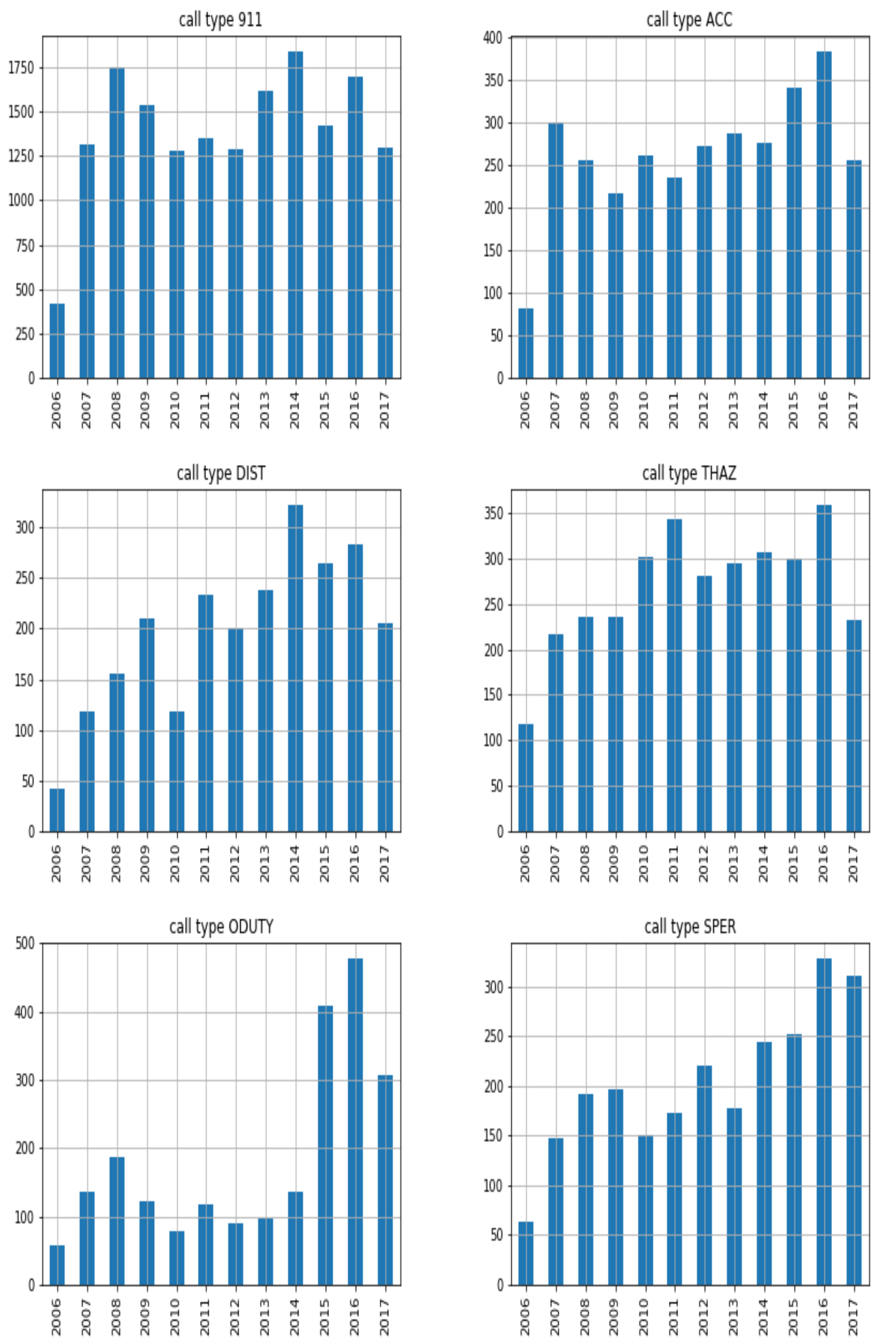


Fig. 24. new search area call types



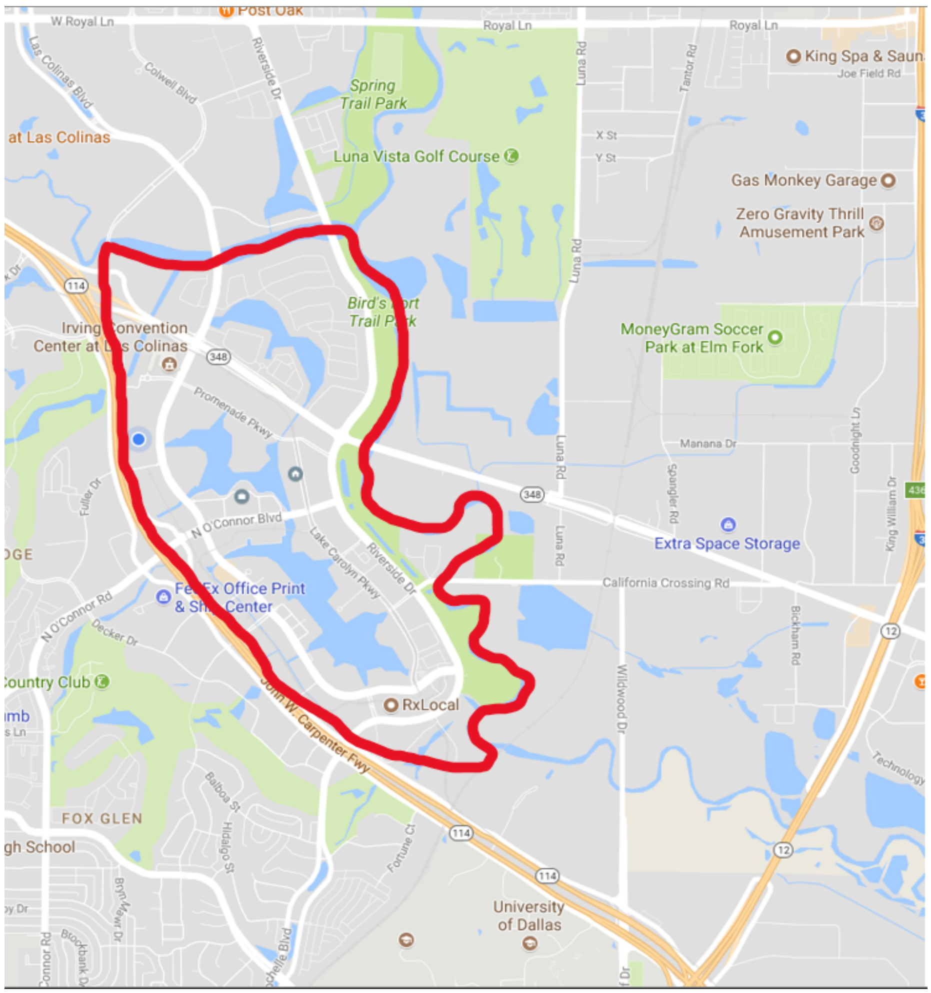


Fig. 25. recommended patrol area